MiniReview

Comparative analyses in aquatic microbial ecology: how far do they go?

Josep M. Gasol a, Carlos M. Duarte b

a Institut de Ciències del Mar, CSIC, Pg. Joan de Borbó s/n, E08039 Barcelona, Spain
b Institut Mediterrani d’Estudis Avançats, CSIC, Universitat de les Illes Balears, Carretera de Valldemossa km 7.5, E07071 Palma de Mallorca, Spain

Received 1 July 1999; received in revised form 18 October 1999; accepted 18 October 1999

Abstract

Methodological developments in recent years have led to an increase in empirical databases on the abundance and functions of aquatic microbes, now allowing synthesis studies. Most of these studies have adopted a comparative approach, such that comparative analyses are now available for most aspects of aquatic microbial food webs (more than 50 papers published in the last 15 years). Some of these analyses apparently yield conflicting results, introducing confusion and unnecessary disputes in the field. We briefly review the comparative analyses so far produced and we highlight generalities, show that some of the perceived discrepancies largely derive from partial analyses of a general underlying trend and formulate predictions based on these general trends that provide new avenues for research. © 2000 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: Comparative analysis; Model; Bacterium; Protist; Virus; Microbial ecology; Ecosystem metabolism; Ecosystem structure

1. Introduction

Microbes are key components of the structure and function of aquatic ecosystems: bacteria, pico- and nanoalgae, protozoa and viruses generally contribute a dominant fraction of the biomass of planktonic ecosystems [1] and their activity often dominates ecosystem functions [2]. Increasing awareness of the role of planktonic microbes has led to a rapid growth of descriptive and experimental studies of their abundance, biomass and activity. This growth has been made possible by the development of robust standardized methods to quantify the abundance of marine microbes and the processes they control [3,4]. The circumvention of methodological problems fueled an avalanche, ongoing since the mid 80’s, of descriptive studies providing estimates of microbial abundance and activity. Now, when the ranges of abundances and activities of aquatic microbes are well constrained, the challenge for microbial ecology for the coming decade relies on the assimilation of this wealth of information into theories with predictive power [4,5].

Examination of recent efforts to achieve such syntheses, which include ecosystem mechanistic models, conceptual models and comparative analyses, has identified the last as a particularly useful approach [4]. The usefulness of pioneer comparative analyses in aquatic microbial ecology published about a decade ago [4] prompted a rapid development of this approach in the past few years. These comparative analyses have been dominated by bivariate correlational studies. Yet, present understanding of the controls on microbial abundance and activity highlights their complexity, involving interactions in the food web and multiple resources, of which bivariate analyses can provide only a fragmented view [6,7]. Thus, whether a coherent, internally consistent body of theory can emerge from the concerted consideration of the piece-meal limited comparative analyses developed to date is not warranted.

Here we provide an overview of the procedures, recent achievements and limitations of comparative analyses in aquatic microbial ecology with the aim of testing the internal consistency of the results derived. We identify apparent conflicts and limitations and suggest the likely routes for progress in the near future.

2. What are comparative analyses?

Comparative analyses are statistical models that draw from previous knowledge to define the range of possible values a distinct process or attribute may adopt, and describe patterns – or the most probable values – in its variability among systems. They do so through the statis-
tical analysis of assemblages of data obtained across systems or derived from the published literature, frequently using existing gradients or contrasts as natural experiments to identify the patterns [6,8]. Typically, but not necessarily, they focus on simple, well-defined variables (e.g. temperature), aggregate properties (e.g. chlorophyll) and relatively large systems (e.g. ecosystems). Although the tools available for comparative analyses are many [7,9], their application in microbial ecology is dominated by bivariate statistical relationships described through regression analysis (see, as an example, Table 1).

A few hundred papers on aquatic microbial ecology are published every year. The capacity to use this wealth of information depends on the availability of efficient approaches to assimilate and synthesize it. Mechanistic models use only a limited amount of empirically determined values. In contrast, the power of comparative analyses relies on the volume of the data used and they are, therefore, best suited to making use of the wealth of available information. By placing the individually obtained data in a broader context, comparative analyses confer added value to the efforts of the individual scientists that have produced the descriptive and experimental work (85% of the papers published, [4]).

Over 50 comparative analyses of aquatic microbial communities have been published to date, half of them over the past 4 years (a full set of references is available through anonymous ftp at ftp.icm.csic.es/pub/gasol). This allows the identification of general procedures in their construction, which resemble those used in similar analyses in general ecology (cf. [7,10]). The formulation of hypotheses and identification of the variables involved is followed by a compilation of the available data (from own sources or from the published literature), which is scrutinized for comparability (i.e. methods used, conversion factors and units) prior to the quantitative analysis of the relationship between the independent and dependent variables [7]. As in general ecology [7], regression analysis is the approach of choice to describe the relationship between the variables, typically through power relationships of the form $Y = ax^b$, common to allometry and fractal analysis [11]. The relationships derived may be used, once validated, to formulate hypotheses or predictions and to draw inferences on the functioning of microbial communities. These can be derived from the slopes of the relationships, their intercepts and the identification of outliers [8]. These inferences, and not the specific predictions from the equations themselves, have provided major insight into the ecology of aquatic microbial communities.

3. Achievements in aquatic microbial ecology

Comparative analyses in aquatic microbial ecology have focused on food web structure, dynamics and function (Table 1). Food web structure has been characterized by analyses involving the abundance, biomass and size of the microbes; food web dynamics by characterizing activity, production and growth rates of the microbes; and their consequences on ecosystem function have been examined by searching for patterns in ecosystem biomass structure and ecosystem metabolism. These traits are generally hypothesized to vary as a function of temperature, nutrient availability and the abundance and production of primary producers. These variables, together with the abundance and activity of microbes are used as independent (i.e. explanatory) variables in comparative analyses. Most (70 out of 170) different equations derived from such comparative analyses used chlorophyll $a$ concentration as the independent variable, while 30 equations used bacterial abundance and 25 used total phosphorus concentration. Temperature was used in only six of the relationships.

Since the first models describing the variability in bacterial abundance as a function of the chlorophyll $a$ concentration (as a surrogate of resource availability) were formulated in the early eighties [12,13], over 30 similar
analyses have been published, expanding the originally very limited marine and freshwater databases to include rivers [14], salt lakes [15], hypertrophic ecosystems [16] and open ocean communities [17,18]. As a result, the link between bacterial abundance and phytoplankton biomass remains as one of the few undisputed patterns in aquatic microbial ecology. Yet, the published relationships differ widely among analyses, which has fueled some controversy. The power slope between bacterial abundance and chlorophyll a concentration has been revised downwards from initial values of about 0.8 [13] to more recent values of about 0.3 [18] in open ocean communities, indicative of a much slower response of bacterial abundance to increasing phytoplankton biomass, at least in low chlorophyll systems. The differences reflect the contrasting nature of the systems included in the data sets used, for these do not represent a random sample of the existing ecosystems in any given analysis. Yet, a closer examination of these relationships shows that they conform to a well-defined space of probable bacterial abundances (Fig. 1) increasing as the average 0.47 power of the chlorophyll a concentration (Table 1). This increase, particularly well constrained at low phytoplankton biomass, where resource supply may be a greater constraint, indicates that bacterial abundance is highest, relative to phytoplankton biomass, in phytoplankton-poor systems, and that bacterial abundance increases along gradients of productivity more slowly than phytoplankton biomass. Indeed, the examination of the 33 published relationships between bacterial abundance and chlorophyll a concentration provides little support for the existence of systematic differences in these relationships among systems, although the range of the variables used to obtain the relationships varies between different systems (Fig. 1).

The abundance of free viruses has also been observed to increase with increasing chlorophyll a concentration and with increasing bacterial abundance [19]. Yet, the difficulties in distinguishing viruses that affect bacteria from those that affect algae have prevented the functional interpretation of these correlations, although differences in the ratios between free viruses and bacteria with varying trophic and ecosystem type have been suggested [19]. Weimbs and Peduzzi [20] computed maximum viral and flagellate concentrations expected at different bacterial abundances to hypothesize that viral-induced bacterial mortality should be more relevant in rich environments, while flagellate-induced bacterial mortality should be more important in poorer environments.

The relationship between bacterial abundance and bacterial predator abundance has also been exhaustively explored, although some analyses report close relationships [21,22] while others report these relationships to be very weak [23]. These discrepancies were shown to depend on the scale of the analyses. At small temporal scales (i.e. the ‘sample’ scale) the predators on protozoa influenced the relationship between protozoans and bacteria, while at large temporal and spatial scales (i.e. at the ‘ecosystem’ scale) the protozoan predators were also related to the trophy of the systems and did not affect the relationship between bacteria and heterotrophic protozoans [24].

Most of the relationships between the abundance and the biomass of aquatic heterotrophs and those of autotrophs are characterized by power slopes significantly < 1 (Table 1) [1]. This means that even though increasing chlorophyll concentrations are accompanied by increasing bacterial, viral, protozoan and zooplankton biomasses, these increases are smaller than those of autotrophs. A corollary that follows from these relationships is that bacteria are dominant components of the biomass of the open ocean [25] on an areal basis and that inverted biomass pyramids shift to upright pyramids when nutrients become available, an inference that was validated through the comparative analysis of planktonic food web structure [1]. Furthermore, the comparative analyses showed that, on an areal basis, marine systems sustain more heterotrophic biomass for a given autotrophic biomass than limnetic systems. These are, again, hypotheses amenable to experimental or empirical tests.

Comparative analyses allow the formulation of useful predictions which are often used in ecosystem models, or are used to validate or compare methods. For instance, the models that relate protozoan growth rates with size and temperature are often used as a reference in field studies (i.e. [26]), to derive estimates when these measurements are lacking, or as a source of parameters for mechanistic models. Methodological differences have been identified, through comparative analyses, as a source of variability in estimates of flagellate grazing rates on bacteria [27], or in estimates of viral abundance [19].

Most importantly, the interconnection between comparative analyses is now providing a consistent body of theory to build a paradigm of the functioning of microbial
food webs of general application. As an example, the comparative analysis of bacterial production by Cole et al. [28] served to validate the different methods used to estimate bacterial production, showed that bacterial production also increases more slowly (i.e., power slope < 1) than system chlorophyll a concentration but faster than their abundance, and demonstrated that bacterial production is, on average, about 25% of primary production. This highly cited piece of research [4] helped to highlight the importance of bacterial production in aquatic ecosystems. The importance of bacteria in the microbial food web depends, however, on the efficiency by which bacteria convert the carbon acquired for production. Comparative analyses have now shown that this efficiency varies across ecosystems, and with bacterial production and substrate type [29]. Bacteria thriving in productive environments use algal-derived carbon more efficiently, accounting for the faster increase in bacterial production with increasing bacterial abundance. A low bacterial growth efficiency in oligotrophic, nutrient-poor plankton environments indicates that bacterial respiration is proportionally more important than bacterial biomass production in these systems, where bacterial and heterotrophic biomass is also relatively high. High heterotroph biomass, together with high bacterial respiration rates in oligotrophic systems suggest that bacterial respiration should exceed primary production [2], rendering oligotrophic systems net heterotrophic. This prediction has strong implications for whole ecosystem metabolism [2] and requires temporally or spatially external carbon subsidies to planktonic communities to balance carbon budgets. These results have been confirmed by subsequent analyses [30], although whether they apply to the oligotrophic ocean is still controversial [31].

Hence, comparative analyses provide statistically sound predictions and allow the test of conceptual and methodological hypotheses at general levels. Most importantly, combination of the results from comparative analyses is providing new views on the role of microbial communities in the function of aquatic ecosystems that await rigorous test.

4. Limitations of the comparative approaches

4.1. Tractability of the concepts

A major limitation in comparative analyses is the slow pace at which new concepts become amenable to this approach. Important novel concepts are often characterized by a poor capacity to quantify the processes involved. An example of this problem is represented by detritus, where the lack of an accepted method to quantify its concentration (but see [32]) has impaired the acquisition of data and the development of models testing its role and importance. In other cases, the variables used are poor approximations of the state variables or processes they are supposed to represent. Chlorophyll a concentration is often used as surrogate of resource availability to heterotrophs, whereas chlorophyll a concentration represents only a crude approximation of algal biomass, which can also vary systematically [32]. Moreover, it is increasingly evident that a substantial, if not major, portion of the planktonic bacterial metabolism, growth and recycling is carried out by a portion of highly active cells [33], and that models that predict the variability in the numbers of highly active bacteria with increasing trophic status suffer significantly when only those highly active bacteria are accounted for [34]. The relationship between nutrient supply and bacterial abundance improves when only the active (or live) cells are included in the analysis [35]. In fact, the 0.5 power slope describing the scaling between bacterial abundance and chlorophyll a concentration could also result from systematic changes in the proportion of active bacteria with increasing nutrient levels. Similarly, the relationships between the abundance of bacteria and that of their predators may change if only the potentially edible bacterial cells [36] or those bacteria belonging to particular phylogenetic group are considered.

The complexity of food web interactions also poses a challenge to their tractability by comparative analyses. In practice, these difficulties have led to a dominance of bottom-up approaches, compared to attempts to quantify top-down relationships [24]. The dominant use of linear regression is quite limiting for describing and constraining ecological variability. Other statistical tools should be used to allow the comparative analyses of complex relationships [9]. Moreover, the goal of comparative analyses should not only be the prediction of mean values of the traits of interest, but they can be successfully used to define boundaries or constraints to the possible set of values (e.g. [37]).

4.2. The random sample problem

A source of concern is the variability among relationships derived from different data sets (Fig. 1). The variability among regression coefficients is remarkably high among studies (Table 1), challenging the belief that, by assembling data from widely diverse ecosystems, the resulting patterns would be of general applicability. That this is not the case is a result of the fact that sampling theory would require the data assembled to represent a random sample of the ecosystems addressed by the comparative analysis (e.g. all lakes, all rivers, the ocean). Such an ideal situation is obviously impossible, for the locations of the ecosystems investigated are unavoidably clustered around major research laboratories and fashionable areas of the ocean to study (e.g. the Sargasso Sea, the Equatorial Pacific). This limitation can, however, be surpassed once the data sets and comparative analyses become so large as to allow the emergence of a common pattern from the seemingly diverse relationships (Fig. 1). Argu-
ments about differences in regression slopes and other properties of the relationships may be, therefore, equivalent to the arguments between blind men trying to identify an elephant by touching different parts, unless the empirical basis for the relationships is so substantial as to suffer little change from the investigation of additional ecosystems.

4.3. Limited precision and rules of thumb

Most of the relationships obtained have order of magnitude resolution and considerable scatter about the mean expectations, leading to poor predictive capacity, particularly when the interest lies in the behavior of individual ecosystems (e.g. Fig. 1). Methodological differences between the researchers originally generating the data (experimental procedures, conversion factors, etc.) are an important source of variability, making the use of these relationships for predictive purposes pretty much useless. The strength of the relationships derived from the comparative analyses lies, thus, on their capacity to uncover general rules and patterns, and to serve as useful 'rules of thumb' rather than their capacity to yield precise predictions. Examples are the expected number of viruses per bacterium [19], the ratio of bacterial production to primary production [28] or the average number of bacteria eaten by a flagellate [27], values that are commonly used as null hypotheses against which newly obtained values are compared.

4.4. Other problems

Most, if not all, of the comparative analyses published to date rely on bivariate analyses, while they address relationships that are intertwined within the complex dynamics of microbial food webs. That the concerted consideration of these piece-meal relationships conform to a coherent body of theory is, therefore, not warranted. Major stumbling blocks for the achievement of internal consistency are provided by (i) possible spurious relationships, derived from the relationship of both variables in a comparative analysis to a third variable not included there; and (ii) lack of consideration of possible time lags between the variables. The possibility that some of the relationships derived in comparative analyses may be spurious has not received sufficient attention as yet. The consideration of the importance of possible time lags has been used to explain the negative relationships between bacterial abundance and chlorophyll a observed in time series as a result of time lags, involving the transference of phytoplankton carbon to DOC following bloom collapses, on a functionally positive relationship [38]. Yet, the confounding effect of such time lags is overridden by the use of large scale comparisons (e.g. Fig. 1), confirming the dependence of the sign of predator-prey or resource-consumer relationships on the scale of the analysis [39].

5. Consistency of current knowledge and potentials of the comparative approaches

The internal consistency of comparative analyses may be addressed through statistical methods, such as path analyses [9], or, better yet, be challenged by the examination of their external consistency, that is, the use of comparative analyses to yield predictions that can be tested against established theories or observations. As an example, the slopes describing the scaling of heterotrophic biomass to that of autotrophs are always \(< 1\) (Table 1). This indicates that heterotroph biomass increases more slowly than that of autotrophs. These relationships also predict the biomass of heterotrophs to be higher than that of autotrophs, in unproductive ecosystems. These observations allowed the formulation of the hypothesis that the biomass pyramid in planktonic environments shifts from a dominance of heterotrophs to a dominance of autotrophs with increasing nutrient levels (Fig. 2), as confirmed by comparative analyses [1], and also through experimental manipulations of planktonic communities. A shift from a dominance of heterotrophic to autotrophic biomass with increasing production suggested that there could be a parallel pattern in community metabolism, where respiratory (biolytic) processes should be dominant over productive (biogenic) processes in the most oligotrophic planktonic environments. This was also supported by comparative analyses [2,30]. A minimum of three additional questions on the functioning of oligotrophic ecosystems surface from these analyses (Fig. 2): (i) how are heterotrophs maintained in oligotrophic systems where both autotrophic biomass and production are relatively low?, (ii) given their greater abundance, do heterotrophs regulate autotrophic biomass in oligotrophic systems? and (iii) are the turnover rates of primary producers much higher in nutrient-poor systems than in richer systems? Or, in other words, are heterotrophs in the open oligotrophic ocean fundamentally independent of the rates of contemporaneous primary production? These questions address apparent paradoxes in the functioning of oligotrophic ecosystems that are currently inspiring experimental and comparative research.

The power scaling between components of microbial food webs (Table 1) also provides a simplified model of its functioning (Fig. 3) that can be challenged by existing observations and models. A power slope of 0.5, like that relating bacteria to chlorophyll a concentration, indicates that the relative importance of bacteria decreases with increasing chlorophyll a concentration. A power slope of 1, like that relating viral abundance to bacterial abundance, indicates that the relative importance of viruses does not change systematically with varying bacterial concentration. A power slope of 1.2, like that relating bacterial production to bacterial abundance, indicates that the relative bacterial production (i.e. specific growth rate) tends to increase slowly with increasing bacterial abundance.
Slopes of biomass heterotrophs vs. biomass autotrophs are < 1

The ratio $H/A$ is > 1 in oligotrophy

There are equivalent patterns in the ratio Respiration : Production

Are heterotrophs maintained by higher autotroph's growth rates in oligotrophy?

Do heterotrophs regulate autotrophs in oligotrophy?

Are heterotrophs fundamentally independent of contemporaneous primary production in the open ocean?

What sustains heterotrophic metabolism in oligotrophic oceans?

Fig. 2. Logical implications that derive from a given pattern identified through comparative analysis. The hypotheses in italics have not yet been subject to analysis. The ratio $H/A$ is the ratio between total heterotrophic biomass and total autotrophic biomass.

Hence, primary production increases proportionally with increasing autotroph biomass, whereas the importance of bacteria declines with increasing chlorophyll levels (Fig. 3A). Viral abundance changes proportionally to bacterial abundance across the productivity gradient while the abundance of heterotrophic nanoflagellates declines with increasing bacterial concentration (Fig. 3A). This last pattern is exactly as predicted by Weimbauer and Peduzzi [20]: heterotrophic nanoflagellates are expected to be more important as sources of bacterial mortality in oligotrophic systems while viruses should be more important in richer environments.

We expect bacterial abundance to increase with larger nutrient supply. While bacterial production does so in a roughly constant way, bacterial abundance does not increase proportionally with chlorophyll. This systematic relative decline in bacterial abundance with increasing chlorophyll a concentration demonstrated by comparative analyses can be interpreted to indicate that bacterial levels are too low for the available resources at high chlorophyll a concentration or, conversely, that bacterial abundance in unproductive ecosystems is too high for the available resources (Fig. 3B). The first explanation (case I in Fig. 3B) suggests that the availability of primary production to bacteria may be reduced by processes like increased sedimentation or reduced carbon exudation by the algae, and/or that bacteria experience a greater grazing pressure in these systems. The hypotheses of an excess bacterial abundance relative to primary production in oligotrophic eco-

systems (case II in Fig. 3B) suggest that bacteria there are fueled by allochthonous carbon subsidies [2] or depend directly on upwelled nutrients which are not processed by algae. Because the concept of 'available resource level' is non-operational, there are serious difficulties for defining a way to discriminate which of the situations is the one that happens in nature.

The examination of the external consistency of comparative analysis in aquatic microbial ecology, such as the analysis presented in Fig. 2, provides, in addition, an effective mechanism of hypothesis generating that enhances the capabilities of the comparative analyses. Indeed, the potential of this research program lies in the inferences made from the equations, and not in the specific predictions from the equations themselves. The analysis presented in Fig. 3 demonstrates the power of comparative analyses to generate, when considered in concert, hypotheses conforming to a coherent view of the functioning of microbial communities across gradients of productivity. The challenge is the elucidation of the questions inspired by these analyses to develop a robust body of theory on the functioning of microbial communities with predictive power. The wealth of information generated over the past decade has demonstrated that the functioning of aquatic microbial communities dominates the metabolism of most aquatic ecosystems, including the global ocean. The challenges ahead pave the way for comparative analyses of microbial communities to provide a key contribution to our understanding of the dynamics of the biosphere.
Fig. 3. A: A simplified model of a microbial food web. The fluxes have values of the average log-log slopes encountered in the different comparative analyses of microbial food webs. The graphs on the sides represent the relative change of a given variable with the increase in the $X$ variable, and are a different way of looking at the meaning of the log-log slopes. B: Two alternative possibilities that are possible when the log-log slope that relates bacteria with chlorophyll is lower than 1. Because the line labeled ‘Maximal use of resources’ is impossible to compute, we do not know if the measured relationship is below (Case I) or above that line (Case II) and, thus, the abundance is maximal or minimal for the available resources. BA, bacterial abundance; VA, viral abundance; BP, bacterial production; CHL, chlorophyll; PP, primary production; SGR, bacterial specific growth rate; HNF, flagellate abundance.

Acknowledgements

We thank F. Thingstad and B. Riemann for the invitation to participate in the MAST Advanced Study Course ‘Concepts and models of the microbial Food Web’ which originated this paper. S. Canut helped in the usual ways. Comments by X.A.G. Morán, C. Pedrós-Alió, M.L. Pace and T.F. Thingstad are greatly appreciated. The work of the authors has been supported by projects: MAS3-CT95-0016 (MEDEA), MAR95-1901-C03-03 (MIDAS) and MAS3-CT96-0045 (COMWEB), funded by the MAST program of the European Commission.
References


